On weak lumpability in Markov chains

Gerardo Rubino, Bruno Sericola

▶ To cite this version:

Gerardo Rubino, Bruno Sericola. On weak lumpability in Markov chains. [Research Report] RR-0787, INRIA. 1987. inria-00075764

HAL Id: inria-00075764
https://inria.hal.science/inria-00075764

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 787

ON WEAK LUMPABILITY IN MARKOV CHAINS

Gerardo RUBINO
Bruno SERICOLA

DECEMBRE 1987

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 36 20 00
Télex: UNIRISA 950 473 F
Télécopie : 99 38 38 32

On Weak Lumpability in Markov Chains

Gerardo Rubino and Bruno Sericola

Décembre 1987, 16 pages

Publication Interne n° 387 - 16 pages - Décembre 87.

Abstract We analyse under which conditions the aggregated process constructed from an homogeneous Markov chain over a given partition of the state space is also Markov homogeneous. The past work on the subject is revised and new properties are obtained.

MARKOV CHAINS, AGGREGATION, LUMPABILITY, WEAK LUMPABILITY

Sur l'Agrégation Faible dans les Chaînes de Markov

Résumé Nous analysons les conditions sous lesquelles le processus agrégé construit à partir d'une chaîne de Markov homogène, relativement à une partition de l'espace d'état, est aussi Markovien homogène. Les travaux passés sur ce sujet sont revus et de nouvelles propriétés sont obtenues.

CHAINES DE MARKOV, AGREGATION, AGREGATION FAIBLE

PAPIER RECUPERE ET RECYCLE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L. A. 227)
UNIVERSITÉ DE RENNES 1 I. N. S. A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

1 Introduction

Let $X = (X_n)_{n\geq 0}$ be an homogeneous irreducible Markov chain evolving in discrete time. To simplify the presentation, the state space is supposed finite and it is denoted $E = \{1, 2, ..., N\}$. The stationary distribution of X is denoted π. Let us denote $\mathcal{B} = \{B(1), B(2), ..., B(M)\}$ a partition of the state space and $n(i)$ the cardinal of $B(i)$. We suppose the states of E ordered such that:

$$\begin{aligned} B(1) &= \{1, \ldots, n(1)\} \\ &\vdots \\ B(m) &= \{n(1) + \ ... + n(m-1) + 1, \ldots, n(1) + \ ... + n(m)\} \\ &\vdots \\ B(M) &= \{n(1) + \ ... \ + n(M-1) + 1, \ldots, N\} \end{aligned}$$

To the given process X we associate the aggregated stochastic process Y with values on $F = \{1, 2, ..., M\}$, defined by:

$$Y_n = m \iff X_n \in B(m) \ \ \forall\, n \geq 0$$

It is easily checked from this definition and the irreducibility of X that the process Y is also irreducible. Y need not be Markov, not even homogeneous. The question here is under which conditions Y is another homogeneous Markov chain.

X is given by its transition probability matrix P and its initial distribution α; we will denote (α, P) this Markov chain when necessary. The elements of P are denoted $P(i, j)$. We will denote $agg(\alpha, P, \mathcal{B})$ the aggregated chain constructed from (α, P) over $\mathcal{B}$. We will often consider the family of all the homogeneous Markov chains over the same state space E sharing the same transition probability matrix P, which can be denoted $(\,.\,, P)$.

Let us denote $\mathcal{A}$ the set of all probability vectors with N entries. It is well known that $agg(\alpha, P, \mathcal{B})$ is an homogeneous Markov chain $\forall\, \alpha \in \mathcal{A}$ if and only if a condition (see next section) is satisfied by the matrix P. Such a chain is called *lumpable* or *strongly lumpable* with repect to $\mathcal{B}$. A more general problem is to determine if there exists some initial distributions α such that $agg(\alpha, P, \mathcal{B})$ is an homogeneous Markov chain but not necessarily for every vector of $\mathcal{A}$. In this case, the Markov chain X is called *weakly lumpable* with respect to the partition $\mathcal{B}$. This problem has been discussed in [1] when X is regular. The authors show that it is possible to have such a situation, that is, $agg(\alpha, P, \mathcal{B})$ Markov homogeneous for some α but not for every starting vector; a simple but strong sufficient condition is also given.

In [2] this analysis is continued by working on a characterization of weak lumpability. A useful technique is introduced by the authors but they give a wrong theorem, and their supposed characterization can only be seen as another sufficient condition. In this paper we will review these works and in particular a counterexample to the main result of [2] will be exibited. The correct form of the condition will be then exposed and we will show that it is a particular case of a more general sufficient condition for weak lumpability. A characterization and some properties of the set of all the initial distributions leading to an homogeneous aggregated Markov chain are given. These results are valid in the more general case where X is only irreducible.

2 Preliminaries

Following the book by Kemeny and Snell, we will denote α^B the vector of $\mathcal{A}$ defined by: $\alpha^B(i) = \alpha(i)/K$ iff i belongs to the subset B of the state space E, where $K = \sum_{j \in B} \alpha(j)$, for all α and B such that $K \neq 0$. On the other side, we will denote α_B the restriction of α to the subset B. We will always consider subsets B belonging to $\mathcal{B}$; so, the elements of B are consecutive integers, and thus, if $B = \{j+1, \ldots, j+k\}$ then $\alpha_B(i) = \alpha(j+i)$, $i = 1, 2, \ldots, k$. For each $l \in F$ the mapping $\alpha \mapsto \alpha_{B(l)}$ will be denoted T_l, that is, $T_l.\alpha = \alpha_{B(l)}$. Conversely, for any vector β with $n(l)$ entries, $T_l^{-1}.\beta = \gamma \in \mathcal{A}$ with $\gamma(i) = 0$ if $i \notin B(l)$ and $\gamma(i) = \beta(i)$ for all $i \in B(l)$. Let us give some examples. Suppose $N = 5$ and $\mathcal{B} = \{B(1), B(2)\}$ with $B(1) = \{1, 2, 3\}$ and $B(2) = \{4, 5\}$. Then, we have:

for $\alpha = (1/10, 1/10, 1/5, 2/5, 1/5)$:

$$\alpha^{B(1)} = (1/4, 1/4, 1/2, 0, 0) \qquad \alpha^{B(2)} = (0, 0, 0, 2/3, 1/3)$$
$$\alpha_{B(1)} = (1/10, 1/10, 1/5) \qquad \alpha_{B(2)} = (2/5, 1/5)$$

$$T_1.\alpha^{B(1)} = (1/4, 1/4, 1/2)$$
$$T_2^{-1}.(1/4, 3/4) = (0, 0, 0, 1/4, 3/4)$$

for $\alpha = (1/2, 1/2, 0, 0, 0)$

$$\alpha^{B(1)} = (1/2, 1/2, 0, 0, 0) \quad \alpha^{B(2)} \text{is not defined}$$
$$(\alpha_{B(2)} = (0, 0))$$

All vectors used in the text are row vectors. Column vectors will be indicated by means of the operator $(.)^T$. A vector with all its entries equal to 1 will be denoted simply 1; its dimension will be defined by the context.

When we need to fix a particular chain of the family $(\,.\,, P)$ by choosing an initial distribution, it is usual to denote explicitely this choice when calculating probabilities by indexing the symbol associated to the probability measure with the initial distribution. For instance, $P_\beta(X_n \in B)$ means "the probability that the state of the chain (β, P) will be in the subset B of E at step n (E and the transition matrix P are considered here as fixed)".

We will denote $P(i, B)$ the transition probability to pass in one step from the state i to the subset B of E, that is $P(i, B) \stackrel{\text{def}}{=} \sum_{j \in B} P(i, j)$. If we consider the decomposition of matrix P in blocks generated by the partition $\mathcal{B}$, we note $P_{B(i)B(j)}$ the block $n(i) \times n(j)$ corresponding to the transitions from $B(i)$ to $B(j)$.

Let us analyse now the chain $Y = agg(\alpha, P, \mathcal{B})$. We will say that the family $(., P)$ is *lumpable* or *strongly lumpable* with respect to $\mathcal{B}$ if for every initial distribution $\alpha \in \mathcal{A}$, $Y = agg(\alpha, P, \mathcal{B})$ is an homogeneous Markov chain. The following well known result characterises these Markov chains:

Theorem 2.1 ([1]) *$(., P)$ is strongly lumpable with respect to $\mathcal{B}$ if and only if for every pair of sets $D, B \in \mathcal{B}$, the probability $P(d, B)$ has the same value for any $d \in D$. This common value is the transition probability corresponding to the aggregated chain Y of moving from set D to set B.*

Let us introduce now another useful notation. A sequence $(C_0, C_1, \ldots, C_j)$ of subsets of E is called a *possible* sequence for α iff $P_\alpha(X_0 \in C_0, \ldots, X_j \in C_j) > 0$. In particular, if $B \in \mathcal{B}$, (B) is a possible sequence for α if $\alpha_B \neq 0$. In the sequel, every considered sequence of elements of $\mathcal{B}$ will be possible for the corresponding starting distribution.

Given any distribution vector $\alpha \in \mathcal{A}$ and a possible sequence $(C_0, C_1, \ldots, C_j)$, we define the vector $f(\alpha, C_0, C_1, \ldots, C_j) \in \mathcal{A}$ recursively by:

$$\begin{aligned} f(\alpha, C) &\stackrel{\text{def}}{=} \alpha^C \\ f(\alpha, C_0, C_1, \ldots, C_k) &\stackrel{\text{def}}{=} (f(\alpha, C_0, C_1, \ldots, C_{k-1})P)^{C_k} \end{aligned}$$

We will denote $\mathcal{A}(\alpha, B)$, for any $B \in \mathcal{B}$, the subset of all the probability distributions of the form $f(\alpha, C_1, \ldots, B)$, that is:

$$\begin{aligned} \mathcal{A}(\alpha, B) \stackrel{\text{def}}{=} \{\ \beta \in \mathcal{A}\ /\ \exists j \geq 0 \text{ and a sequence } (C_1,\ \ldots, C_j), \text{ empty if } j = 0, \\ \text{such that } \beta = f(\alpha,\ C_1,\ \ldots, C_j,\ B)\ \} \end{aligned}$$

It is easy to verify that for all $\alpha \in \mathcal{A}$ and $B \in \mathcal{B}$, the family $\mathcal{A}(\alpha, B)$ is not empty (consequence of the irreducibility of Y). With this notation, a first characterization of the fact that Y is an homogeneous Markov chain is given by the following theorem:

Theorem 2.2 (Kemeny-Snell [1]) *The chain $Y = agg(\alpha, P, \mathcal{B})$ is an homogeneous Markov chain iff $\forall B, D \in \mathcal{B}$, the probability $P_\beta(X_1 \in B)$ is the same for every $\beta \in \mathcal{A}(\alpha, D)$. This common value is the transition probability corresponding to the aggregated chain Y of moving from set D to set B.*

Kemeny and Snell proove in [1] the unicity of the transition probability matrix of the aggregated homogeneous Markov chain when the chain X is regular ($\mathcal{B}$ is fixed):

Corollary 2.3 (**[1]**) *If $\hat{P}$ is the transition probability matrix of the aggregated homogeneous Markov chain $Y = agg(\alpha, P, \mathcal{B})$ then $\hat{P}$ is the same for every α leading to an aggregated homogeneous Markov chain.*

We will see that if X is only irreducible, then Corollary 2.3 is still valid (Theorem 3.5 in the next section).

The set of initial distributions α leading to an homogeneous Markov chain for $Y = agg(\alpha, P, \mathcal{B})$ will be denoted $\mathcal{A}_\mathcal{M}$, that is:

$$\mathcal{A}_\mathcal{M} \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A} / Y = agg(\alpha, P, \mathcal{B}) \text{ is an homogeneous Markov chain}\}$$

The case $\mathcal{A}_\mathcal{M} \neq \emptyset$ and $\mathcal{A}_\mathcal{M} \neq \mathcal{A}$ exists as shown in [1]. It is also shown that when the chain X is regular, if $\mathcal{A}_\mathcal{M} \neq \emptyset$ then $\pi \in \mathcal{A}_\mathcal{M}$. This is a consequence of the fact that if $\alpha \in \mathcal{A}_\mathcal{M}$ then $\alpha P \in \mathcal{A}_\mathcal{M}$.

Let us consider now the following definition and notation:

- We shall say that $(.,P)$ is ***weakly lumpable*** with respect to the partition $\mathcal{B}$ iff $\mathcal{A}_{\mathcal{M}} \neq \emptyset$. For any $\alpha \in \mathcal{A}_{\mathcal{M}}$, the aggregated chain $Y = agg(\alpha, P, \mathcal{B})$ is then an homogeneous Markov chain and we shall denote by $\hat{P}$ its transition probability matrix which is the same for every $\alpha \in \mathcal{A}_{\mathcal{M}}$ (Corollary 2.3).

- For $i \in F$, we shall denote by $\tilde{P}_i$ the $(n(i), M)$ matrix with:
$\tilde{P}_i(j,k) = P(\, n(1) + \cdots + n(i-1) + j, B(k)\,),\, 1 \leq j \leq n(i),\, k \in F$

From Corollary 2.3, we deduce easily the relation: $\hat{P}_j = (T_j.\pi^{B(j)})\tilde{P}_j$, where $\hat{P}_j$ denotes the j^{th} row of $\hat{P}$. In the sequel, we shall always define $\hat{P}$ according to the previous expression even if the aggregated chain is not an homogeneous Markov chain. That is, if Y is Markov homogeneous, $\hat{P}$ is its transition probability matrix; in the other cases, $\hat{P}_j \stackrel{\text{def}}{=} (T_j.\pi^{B(j)})\tilde{P}_j$, $\forall j \in F$

- For $j \in F$, we shall denote by σ_j the following linear system in the vector x_j: $x_j\tilde{P}_j = \hat{P}_j$. It can be seen as a system in M equations and $n(j)$ scalar unknowns. It is easy to verify that if x_j is a solution to σ_j then $x_j 1^T = 1$. By construction, this system has, at least, the solution $x_j = T_j.\pi^{B(j)}$.

- $\mathcal{A}^* \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A} \,/\, T_l.\alpha^{B(l)}$ is solution of σ_l for every $l \in F$ such that $\alpha_{B(l)} \neq 0 \,\}$.

Remark that $\mathcal{A}^* \neq \emptyset$ because by construction, $\pi \in \mathcal{A}^*$.

- We shall say that a subset $\mathcal{U}$ of $\mathcal{A}$ is ***stable by right product by*** P iff $\forall x \in \mathcal{U}$ the vector $xP \in \mathcal{U}$.

Now, in [2] the authors give the following characterization of weak lumpability:

"$\mathcal{A}_{\mathcal{M}} \neq \emptyset \Longleftrightarrow \mathcal{A}^*$ is stable by right product by P. In this case, $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^*$".

In fact, the condition proposed to characterize weak lumpability has, in their paper, the less compact following form [2, Theorem 3.1]:

$\forall l \in F$, [y_l is a solution to $\sigma_l \Longrightarrow \forall m \in F$ such that $((T_l^{-1}.y_l)P)_{B(m)} \neq 0$,
$T_m.((T_l^{-1}.y_l)P)^{B(m)}$ is a solution to σ_m].

In the Appendix we show that this is equivalent to the stability of $\mathcal{A}^*$ by right product by P.

This theorem is not true as it is shown in the following counterexample:

$$P = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/2 \\ 1/8 & 3/8 & 1/4 & 1/4 \\ 3/8 & 1/8 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

and $\mathcal{B} = \{B(1), B(2)\}$, where $B(1) = \{1,2,3\}$ and $B(2) = \{4\}$.
We have:

$$\pi = (3/13, 3/13, 3/13, 4/13)$$

Now, check first that

$$\pi^{B(1)} = (1/3, 1/3, 1/3, 0) \text{ and that } \pi^{B(1)}P = (2/9, 2/9, 2/9, 1/3)$$

Next, remark that $(\pi^{B(i)}P)^{B(j)} = \pi^{B(j)}$ for every $i, j \in \{1,2\}$. So, it is evident from Theorem 2.2, as it is shown in [1], that this is sufficient for $\pi \in \mathcal{A}_{\mathcal{M}}$, since we will have $\mathcal{A}(\pi, B(i)) = \{\pi^{B(i)}\}$ for $i = 1, 2$. Then:

$$\tilde{P}_1 = \begin{pmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \\ 3/4 & 1/4 \end{pmatrix} \qquad \tilde{P}_2 = \begin{pmatrix} 3/4 & 1/4 \end{pmatrix}$$

which gives:

$$\hat{P} = \begin{pmatrix} 2/3 & 1/3 \\ 3/4 & 1/4 \end{pmatrix}$$

The vector $u = (1/3, 0, 2/3, 0) \in \mathcal{A}^*$ since $T_1.u^{B(1)}$ is a solution to σ_1 ($u^{B(2)}$ is not defined).

Consider now $v = uP = (11/36, 5/36, 2/9, 1/3)$. $v^{B(1)} = (11/24, 5/24, 1/3, 0)$ and $T_1.v^{B(1)}\tilde{P}_1 = (61/96, 35/96) \neq \hat{P}_1$. So, $T_1.v^{B(1)}$ is not a solution to σ_1 and thus $uP \notin \mathcal{A}^*$.

This is an example of an homogeneous Markov chain and a partition of the state space with $\mathcal{A}_{\mathcal{M}} \neq \emptyset$ and at the same time $\mathcal{A}^*$ not stable by right product by P.

Moreover, in [2] the authors derive some properties of $\mathcal{A}_{\mathcal{M}}$ such as the convexity of this set, but they use the preceeding inexact characterization and in fact they only give properties of $\mathcal{A}^*$ (see Section 3 for an analysis of $\mathcal{A}_{\mathcal{M}}$).

3 Main results

All the results given in this section are under the only hypothesis that the chain X is irreducible. We first give the correct form of the previous equivalence and then we exhibit a characterization of the set $\mathcal{A}_{\mathcal{M}}$.

Theorem 3.1 *$\mathcal{A}^*$ is stable by right product by $P \Longleftrightarrow \mathcal{A}_{\mathcal{M}} = \mathcal{A}^*$.*

Proof. $\mathcal{A}_{\mathcal{M}}$ being obviously stable by right product by P, the right to left implication becomes trivial. To proove the converse, let us remark first that Theorem 2.2 can be written as:

$agg(\alpha, P, \mathcal{B})$ is Markov homogeneous $\Longleftrightarrow \forall l \in F,\ T_l.\beta$ is a solution to σ_l for every $\beta \in \mathcal{A}(\alpha, B(l))$.

It follows clearly that $\mathcal{A}_{\mathcal{M}} \subseteq \mathcal{A}^*$. Let then $\alpha \in \mathcal{A}^*$ and consider a possible sequence $(B(i), B(j), \ldots)$ for α. We have:

$\alpha^{B(i)} \in \mathcal{A}^*$ trivially since $(\alpha^{B(i)})^{B(i)} = \alpha^{B(i)}$
$\alpha^{B(i)}P \in \mathcal{A}^*$ by hypothesis
$(\alpha^{B(i)}P)^{B(j)} \in \mathcal{A}^*$ for every $i, j \in F$ by definition of $\mathcal{A}^*$
$(\alpha^{B(i)}P)^{B(j)}P \in \mathcal{A}^*$ for every $i, j \in F$ by hypothesis

By continuing in this way, a recurrence argument gives:

$\mathcal{A}(\alpha, B(l)) \subseteq \mathcal{A}^*$ for every $l \in F$, which is equivalent to say that $agg(\alpha, P, \mathcal{B})$ is Markov homogeneous (Theorem 2.2). So, $\alpha \in \mathcal{A}_{\mathcal{M}}$. $\square$

We see then that the stability of $\mathcal{A}^*$ by right product by P is not a characterization of weak lumpability but only a sufficient condition.

Let us now introduce the following lemmas which will be useful for the sequel.

Lemma 3.2 *The function $v \mapsto P_v(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k})$ for fixed j, k, $C_{j-1}, \ldots, C_{j-k}$ and B, is continuous.*

Proof. Let us denote $g(v) = P_v(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k})$. Then:

$$g(v) = \sum_{i \in E} v(i) P_v(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k}, X_0 = i)$$

But $P_v(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k}, X_0 = i)$ does not depend on v. We can then write:

$$\begin{aligned}
|g(v') - g(v)| &= |\sum_{i \in E} (v'(i) - v(i)) P(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k}, X_0 = i)| \\
&\leq \sum_{i \in E} |v'(i) - v(i)| P(X_j \in B/X_{j-1} \in C_{j-1}, \ldots, X_{j-k} \in C_{j-k}, X_0 = i) \\
&\leq \sum_{i \in E} |v'(i) - v(i)| \\
&= \|v' - v\|
\end{aligned}$$

and the continuity of g follows. □

Lemma 3.3 *The restriction of $\beta = f(\alpha, C_0, C_1, \ldots, C_j)$ to the subset C_j is given by the following expression:*

$$\beta_{C_j} = K^{-1} \alpha_{C_0} P_{C_0 C_1} P_{C_1 C_2} \ldots P_{C_{j-1} C_j}$$

$$\textit{where } K = \alpha_{C_0} P_{C_0 C_1} P_{C_1 C_2} \ldots P_{C_{j-1} C_j} 1^T$$

Proof. The property is obvious for $j = 0$. Let us suppose that the relation holds for any sequence having n subsets and denote $\gamma = f(\alpha, C_0, C_1, \ldots, C_{n-1})$. We have:

$$\gamma_{C_{n-1}} = L^{-1} \alpha_{C_0} P_{C_0 C_1} P_{C_1 C_2} \ldots P_{C_{n-2} C_{n-1}}$$

where $L = \alpha_{C_0} P_{C_0C_1} P_{C_1C_2} \dots P_{C_{n-2}C_{n-1}} 1^T$.

Then, if $\beta = f(\alpha, C_0, C_1, \dots, C_n) = (\gamma P)^{C_n}$, we have:

$$\beta_{C_n} = H^{-1}(\gamma P)_{C_n} \quad \text{with } H = (\gamma P)_{C_n} 1^T$$

But $(\gamma P)_{C_n} = \gamma_{C_{n-1}} P_{C_{n-1}C_n}$ since $\gamma(i) = 0$ if $i \notin C_{n-1}$. So,

$$\begin{aligned} \beta_{C_n} &= H^{-1} \gamma_{C_{n-1}} P_{C_{n-1}C_n} \\ &= H^{-1} L^{-1} \alpha_{C_0} P_{C_0C_1} P_{C_1C_2} \dots P_{C_{n-2}C_{n-1}} P_{C_{n-1}C_n} \end{aligned}$$

and the result follows since K^{-1} is simply a normalizing constant. □

Lemma 3.4 *If $\alpha \in \mathcal{A}_{\mathcal{M}}$ then:*

- *$\alpha^{B(k)} \in \mathcal{A}_{\mathcal{M}}$ for every $k \in F$ such that the sequence $(B(k))$ is possible.*
- $\frac{1}{n} \sum_{k=1}^{n} \alpha P^k \in \mathcal{A}_{\mathcal{M}}$

Proof. Let $\alpha \in \mathcal{A}_{\mathcal{M}}$ and $k \in F$ with $\alpha_{B(k)} \neq 0$. Verify first that:

$$P_{\alpha^{B(k)}}(X_{n+1} \in B(m) / X_n \in B(l),\ X_{n-1} \in C_{n-1},\ \dots, X_0 \in B(k)) = P_\delta(X_1 \in B(m))$$

where $\delta = f(\alpha^{B(k)},\ B(k), \dots, C_{n-1},\ B(l)) = f(\alpha,\ B(k),\ \dots, C_{n-1},\ B(l))$.

So:

$$P_\delta(X_1 \in B(m)) = P_\alpha(X_{n+1} \in B(m) / X_n \in B(l),\ X_{n-1} \in C_{n-1},\ \dots, X_0 \in B(k))$$

which prooves that $\alpha^{B(k)} \in \mathcal{A}_{\mathcal{M}}$.

Let us denote:

$$\gamma = \frac{1}{n} \sum_{k=1}^{n} \alpha P^k$$

We have:

$$P_\gamma(X_{n+1} \in B(m) / X_n \in B(l),\ X_{n-1} \in C_{n-1},\ \dots, X_0 \in C_0) = P_{\gamma'}(X_1 \in B(m))$$

where: $\gamma' = f(\gamma,\ C_0,\ \dots,\ C_{n-1},\ B(l))$.

Let $\alpha'_k = f(\alpha P^k,\, C_0,\, \ldots,\, C_{n-1},\, B(l))$. From Lemma 3.3 we have:

$$\gamma' = \frac{1}{n}\sum_{k=1}^{n}\frac{K_k}{K}\alpha'_k$$

where

$$K_k = (\alpha P^k)_{C_0} P_{C_0 C_1} \ldots P_{C_{n-1} B(l)} 1^T$$

So, we can write:

$$P_{\gamma'}(X_1 \in B(m)) = \frac{1}{n}\sum_{k=1}^{n}\frac{K_k}{K} P_{\alpha'_k}(X_1 \in B(m))$$

But:

$$\begin{aligned} P_{\alpha'_k}(X_1 \in B(m)) &= P_{\alpha P^k}(X_{n+1} \in B(m)/X_n \in B(l),\, \ldots,\, X_0 \in C_0) \\ &= P_\alpha(X_{n+k+1} \in B(m)/X_{n+k} \in B(l),\, \ldots,\, X_k \in C_0) \\ &= P_\alpha(X_1 \in B(m)/X_0 \in B(l)) \quad \text{since } \alpha \in \mathcal{A}_\mathcal{M} \end{aligned}$$

So:

$$P_{\gamma'}(X_1 \in B(m)) = P_\alpha(X_1 \in B(m)/X_0 \in B(l))$$

which prooves that $\gamma \in \mathcal{A}_\mathcal{M}$. □

We are now ready to proove the following theorem:

Theorem 3.5 *If X is only irreducible then Corollary 2.3 holds and if $\mathcal{A}_\mathcal{M} \neq \emptyset$ then $\pi \in \mathcal{A}_\mathcal{M}$.*

Proof. Remark first that $\mathcal{A}_\mathcal{M}$ is trivially stable by right product by P without any particular assumption about X such as regularity.

Let $\alpha \in \mathcal{A}$ such that $Y = agg(\alpha, P, \mathcal{B})$ is an homogeneous Markov chain (i.e. $\alpha \in \mathcal{A}_\mathcal{M}$) with transition probability matrix $\hat{P}$. We can write:

$$\begin{aligned} \hat{P}(l,m) &= P_\alpha(X_{k+1} \in B(m)/X_k \in B(l)) \text{ for any } k \geq 0 \\ &= P_{\alpha P^k}(X_1 \in B(m)/X_0 \in B(l)) \text{ for any } k \geq 0 \end{aligned}$$

Now,

$$\alpha \in \mathcal{A}_{\mathcal{M}} \Longrightarrow \frac{1}{n}\sum_{k=1}^{n}\alpha P^k \in \mathcal{A}_{\mathcal{M}} \text{ for every } n \geq 1, \text{ by Lemma 3.4.}$$

To clarify the text, let γ_k denote the quantity;

$$\gamma_k = \frac{(\alpha P^k)_{B(l)}1^T}{\sum_{k=1}^{n}(\alpha P^k)_{B(l)}1^T} \quad \text{for } k = 1, 2, \ldots, n$$

$\hat{P}(l, m)$ can be written as follows:

$$\hat{P}(l,m) = \sum_{k=1}^{n}\gamma_k \hat{P}(l,m) \quad \text{for any } n \geq 1$$

So:

$$\begin{aligned}
\hat{P}(l,m) &= \textstyle\sum_{k=1}^{n}\gamma_k P_{\alpha P^k}(X_1 \in B(m)/X_0 \in B(l)) \quad \text{for any } n \geq 1 \\
&= \textstyle\sum_{k=1}^{n}\gamma_k P_{(\alpha P^k)^{B(l)}}(X_1 \in B(m)) \quad \text{for any } n \geq 1 \\
&= P_{\sum_{k=1}^{n}\gamma_k(\alpha P^k)^{B(l)}}(X_1 \in B(m)) \quad \text{for any } n \geq 1
\end{aligned}$$

The sum in subscript can be written as follows:

$$\sum_{k=1}^{n}\gamma_k(\alpha P^k)^{B(l)} = (\sum_{k=1}^{n}\alpha P^k)^{B(l)} = (\frac{1}{n}\sum_{k=1}^{n}\alpha P^k)^{B(l)} \quad \text{for any } n \geq 1$$

So, we have:

$$\begin{aligned}
\hat{P}(l,m) &= P_{(\frac{1}{n}\sum_{k=1}^{n}\alpha P^k)^{B(l)}}(X_1 \in B(m)) \quad \text{for any } n \geq 1 \\
&= P_{\frac{1}{n}\sum_{k=1}^{n}\alpha P^k}(X_1 \in B(m)/X_0 \in B(l)) \quad \text{for any } n \geq 1
\end{aligned}$$

Letting now $n \longrightarrow +\infty$ and thanks to the continuity Lemma 3.2, we get:

$$\hat{P}(l,m) = P_\pi(X_1 \in B(m)/X_0 \in B(l))$$

which does not depend on α. □

We can now establish the following theorem:

Theorem 3.6 *$\mathcal{A}_{\mathcal{M}}$ is a convex closed set.*

Proof. From the continuity Lemma 3.2, it is immediat that $\mathcal{A}_{\mathcal{M}}$ is closed. To prove the convexity, assume that $\alpha,\ \beta \in \mathcal{A}_{\mathcal{M}}$ and let $\gamma = \lambda\alpha + \mu\beta$, where $\lambda + \mu = 1$.

$$P_{\gamma}(X_{n+1} \in B(m)/X_n \in B(l),\ X_{n-1} \in C_{n-1},\ \ldots, X_0 \in C_0) = P_{\gamma'}(X_1 \in B(m))$$

where $\gamma' = f(\gamma,\ C_0,\ \ldots, C_{n-1},\ B(l))$.

Let $\alpha' = f(\alpha,\ C_0,\ \ldots, C_{n-1},\ B(l))$ and $\beta' = f(\beta,\ C_0,\ \ldots, C_{n-1},\ B(l))$. From Lemma 3.3 we have:

$$\gamma' = \lambda \frac{K_\alpha}{K}\alpha' + \mu \frac{K_\beta}{K}\beta'$$

where
$K_\alpha = \alpha_{C_0} P_{C_0C_1} \ldots P_{C_{j-1}C_j} 1^T$, $K_\beta = \beta_{C_0} P_{C_0C_1} \ldots P_{C_{j-1}C_j} 1^T$
and
$K = \lambda K_\alpha + \mu K_\beta$.

Therefore,

$$\begin{aligned}
P_{\gamma'}(X_1 \in B(m)) &= P_{(\lambda K_\alpha \alpha' + \mu K_\beta \beta')/K}(X_1 \in B(m)) \\
&= \lambda \frac{K_\alpha}{K} P_{\alpha'}(X_1 \in B(m)) + \mu \frac{K_\beta}{K} P_{\beta'}(X_1 \in B(m)) \\
&= \lambda \frac{K_\alpha}{K} \hat{P}(l,m) + \mu \frac{K_\beta}{K} \hat{P}(l,m) \quad \text{by Theorem 3.5} \\
&= \hat{P}(l,m)
\end{aligned}$$

which means that $\gamma \in \mathcal{A}_{\mathcal{M}}$. So $\mathcal{A}_{\mathcal{M}}$ is a convex set. □

To give a characterization of the set $\mathcal{A}_{\mathcal{M}}$, we define for all $j \geq 1$:

$$\mathcal{A}^j \stackrel{\text{def}}{=} \{\alpha \in \mathcal{A}^* \ /\ \forall\beta = f(\alpha,\ B(i_1),\ \ldots, B(i_k)),\ \text{with } k \leq j,\ \beta \in \mathcal{A}^*\}$$

With this notation, we have:

$$\mathcal{A}^1 = \{\alpha \in \mathcal{A}^* \ /\ \forall m \in F \text{ such that } \alpha_{B(m)} \neq 0,\ \alpha^{B(m)} \in \mathcal{A}^*\} = \mathcal{A}^*$$

$(\mathcal{A}^j)_{j\geq 1}$ is obvously a decreasing sequence:

$$\forall j \geq 1,\ \mathcal{A}^{j+1} \subseteq \mathcal{A}^j$$

The following theorem gives then a characterization of the set $\mathcal{A}_{\mathcal{M}}$:

Theorem 3.7

$$\mathcal{A}_{\mathcal{M}} = \bigcap_{j\geq 1} \mathcal{A}^j$$

Proof. As in the proof of Theorem 3.1, we start from the equivalence:

$agg(\alpha, P, \mathcal{B})$ is Markov homogeneous $\Longleftrightarrow$ $\forall l \in F,\ \forall \beta \in \mathcal{A}(\alpha, B(l)),\ T_l.\beta$ is a solution to σ_l.

In other words:

$$\mathcal{A}_{\mathcal{M}} = \{\alpha \in \mathcal{A}^* \ /\ \forall l \in F,\ \mathcal{A}(\alpha, B(l)) \subseteq \mathcal{A}^*\}$$

So, we have: $\forall j \geq 1,\ \mathcal{A}_{\mathcal{M}} \subseteq \mathcal{A}^j$.

Conversely, if $\alpha \in \mathcal{A}^j\ \forall j \geq 1$, then $\forall l \in F\ \mathcal{A}(\alpha, B(l)) \subseteq \mathcal{A}^*$; which means that $\alpha \in \mathcal{A}_{\mathcal{M}}$ □

An evident necessary condition for having weak lumpability (i.e. $\mathcal{A}_{\mathcal{M}} \neq \emptyset$) is that $\forall j \geq 1,\ \mathcal{A}^j \neq \emptyset$

A sufficient condition is given by the following corollary:

Corollary 3.8 *If $\exists j \geq 1$ such that $\mathcal{A}^{j+1} = \mathcal{A}^j$ then $\mathcal{A}_{\mathcal{M}} = \mathcal{A}^{j+k}$, $\forall k \geq 0$.*

Proof. If $\mathcal{A}^j = \emptyset$, the result is trivial.

Let us remark now that:

$$\forall j \geq 1,\ (\alpha \in \mathcal{A}^j \Longrightarrow \alpha^{B(l)} \in \mathcal{A}^j,\ \forall l \in F \text{ such that } \alpha_{B(l)} \neq 0)$$

Then, let $\mathcal{U} = \mathcal{A}^j = \mathcal{A}^{j+1}$ and $\alpha \in \mathcal{U}$. By definition, for any $i_1, \ldots, i_{j+1} \in F$, $f(\alpha, B(i_1), \ldots, B(i_{j+1})) \in \mathcal{A}^*$.

But $f(\alpha,\ B(i_1),\ \ldots, B(i_{j+1})) = f(\alpha^{B(i_1)}P,\ B(i_2),\ \ldots, B(i_{j+1}))$; so, $\alpha^{B(i_1)}P\ \in\ \mathcal{U}$ (since $\mathcal{U} = \mathcal{A}^j$).

Therefore, we have:

$$\alpha \in \mathcal{U} \Longrightarrow \alpha^{B(l)}P \in \mathcal{U},\ \forall l \in F \text{ such that } \alpha_{B(l)} \neq 0$$

This means that if $\alpha \in \mathcal{U}$ then $\forall l \in F,\ \mathcal{A}(\alpha, B(l)) \subseteq \mathcal{U}$ which implies that $\alpha \in \mathcal{A}_{\mathcal{M}}$. Therefore, $\mathcal{U} \subseteq \mathcal{A}_{\mathcal{M}}$. Now, thanks to Theorem 3.7, the result follows. □

Observe that the sufficient condition for weak lumpability of Theorem 3.1 reduces to the preceeding one with $j = 1$.

4 Conclusions

In this paper we analyse the set of initial distributions of an homogeneous Markov chain which lead to an homogenous aggregated Markov chain, given the probability transition matrix and a partition of the state space. We give a characterization and some properties of this set. This characterization is rather theoretical since it involves an infinite product of sets. Further work seems necessary to find more efficient ways to decide whether a given Markov chain is weakly lumpable or not. Also, another direction to explore is the extension of this kind of results to the continuous time case.

References

[1] J.G. Kemeny and J.L Snell. *Finite Markov chains.* Springer-Verlag, New York Heidelberg Berlin, 1976.

[2] A.M. Abdel-Moneim and F.W. Leysieffer. Weak lumpability in finite Markov chains. *J.Appl.Prob.*, 19:685–691, 1982.

Appendix

Lemma. *The two following assertions are equivalent:*

(i) $\mathcal{A}^*$ *is stable by right product by* P

(ii) $\forall l \in F$, [y_l *is a solution to* $\sigma_l \Longrightarrow \forall m \in F$ *such that* $((T_l^{-1}.y_l)P)_{B(m)} \neq 0$, $T_m.((T_l^{-1}.y_l)P)^{B(m)}$ *is solution to* σ_m].

Proof. Let us show first that $\mathcal{A}^*$ is a convex set. Let α and β be two distributions of $\mathcal{A}^*$ and $\lambda \in [0,1]$. Consider the vector $\gamma = \lambda\alpha + (1-\lambda)\beta$ and let $B(l)$ be any element of the partition such that $\gamma_{B(l)} \neq 0$. If $\alpha_{B(l)} = 0$ then $\gamma^{B(l)} = \beta^{B(l)}$ and $\gamma \in \mathcal{A}^*$; in the same way, if $\beta_{B(l)} = 0$, $\gamma \in \mathcal{A}^*$. In the other case, it is easy to check that:

$$\gamma^{B(l)} = \mu\alpha^{B(l)} + (1-\mu)\beta^{B(l)} \text{ where } \quad \mu = \frac{\lambda\alpha_{B(l)}\mathbf{1}^T}{\lambda\alpha_{B(l)}\mathbf{1}^T + (1-\lambda)\beta_{B(l)}\mathbf{1}^T}$$

Then, as the set of the solutions to σ_l is convex, $T_l.\gamma^{B(l)}$ is a solution to this linear system and the convexity of $\mathcal{A}^*$ follows.

(i) $\Longrightarrow$ (ii)

Let y_l a solution to σ_l and $m \in F$ such that $((T_l^{-1}.y_l)P)_{B(m)} \neq 0$. As $T_l^{-1}.y_l$ is obviously in $\mathcal{A}^*$, $T_m.((T_l^{-1}.y_l)P)^{B(m)} \in \mathcal{A}^*$ since $\mathcal{A}^*$ is stable by right product by P, so we have (ii).

(ii) $\Longrightarrow$ (i)

Let $\alpha \in \mathcal{A}^*$. For any $l \in F$ such that $\alpha_{B(l)} \neq 0$, (ii) says that $\forall m \in F$ with $(\alpha^{B(l)}P)_{B(m)} \neq 0$ we have: $T_m.(\alpha^{B(l)}P)^{B(m)}$ is a solution to σ_m; that is, $\alpha^{B(l)}P \in \mathcal{A}^*$. We can write (in an unique way) α as a convex combination of vectors $\alpha^{B(l)}$ which are all in $\mathcal{A}^*$:

$$\alpha = \sum_{l \in F / \alpha_{B(l)} \neq 0} \lambda_l \alpha^{B(l)} \text{ where } \lambda_l = \alpha_{B(l)}\mathbf{1}^T.$$

Then, the convexity of $\mathcal{A}^*$ ends the proof. □

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique